Feedback Choices and Their Relations to Learning Are Age-Invariant Starting in Middle School: A Secondary Data Analysis

Maria Cutumisu[a,*] and Daniel L. Schwartz[b]

cutumisu@ualberta.ca, danls@stanford.edu

[a]University of Alberta, [b]Stanford Graduate School of Education

*Corresponding author:

Dr. Maria Cutumisu, 6-102 Education North Centre, Department of Educational Psychology, Faculty of Education, University of Alberta, Edmonton, T6G 2G5, Canada, Telephone: 780 492 5211, Fax: 780 492 1318, Email: cutumisu@ualberta.ca.

Maria Cutumisu is an associate professor in the Department of Educational Psychology, at the Faculty of Education, University of Alberta. She received M.Sc. and Ph.D. degrees in Computing Science from the University of Alberta and she trained as a postdoctoral scholar in Learning Sciences at the Stanford Graduate School of Education. Her research interests include game-based assessments, artificial intelligence in games, feedback-seeking, data science, and eye tracking.

Daniel L. Schwartz is the Dean of the Stanford University Graduate School of Education and holds the Nomellini-Olivier Chair in Educational Technology. He is a learning scientist who also spent eight years teaching secondary school in Los Angeles and Kaltag, Alaska. His special niche is the ability to produce novel and effective learning activities that also test basic hypotheses about how people learn.

Abstract

The ability to seek feedback about one's learning is an essential skill for independent learners. Although educational researchers have focused on feedback assigned to students, little research has investigated feedback-seeking, and no previous investigation has explored feedback-seeking in relation to both learning and age. This secondary data analysis study examines the effect of age on choices to seek critical feedback and to revise digital posters as well as on learning outcomes for n = 764 students across six US public schools. Previously reported data from three middle schools were merged with new data from a middle school, a high school, and a community college to sample a larger age range. Students played a digital assessment game, Posterlet, which measured their choices to seek critical feedback and to revise posters. This study examines critical feedback-seeking across age, starting in middle school, and it demonstrates that it is possible to capture and measure learning choices (e.g., behaviors that students employ while learning) using the same assessment, Posterlet, across age. The results revealed that, starting in middle school, (1) age did not influence the frequency of critical feedback-seeking, as students' feedback choices were stable across age; (2) critical feedback-seeking correlated with learning of graphic design principles, poster performance, and students' choices to revise their posters; and, importantly, (3) age did not moderate the relation between critical feedback-seeking and learning behaviors and outcomes across a large developmental sample. Thus, critical feedback-seeking is stable from middle school through adulthood and the benefits of seeking critical feedback have the same impact on learning outcomes regardless of age.

Keywords: choice; feedback; learning; revision; game-based assessment; age

Introduction

The assessment of individuals using a stable test across age constitutes a major research goal, as it may help detect cognitive (e.g., intelligence quotient or IQ) or non-cognitive (e.g., mindfulness) changes through development. A large body of research concerned with measurement invariance has investigated whether (1) underlying constructs are measured by an instrument similarly over groups or over time; and (2) the model underlying a set of test scores is directly comparable across age groups, with the goal to enable inferences regarding the measurement of the same psychological trait among age groups (Bowden, Weiss, Holdnack, & Lloyd, 2006; Droutman, Golub, Oganesyan, & Read, 2018; Dunn & Dunn, 2007; Weintraub et al., 2013). Such age-independent assessments, which are standardized for use regardless of age, enable professionals, educators, parents, and researchers to detect and measure changes (e.g., cognitive shifts within a skill set) across time. This enables the undertaking of early screening that can inform appropriate scaffolding and interventions.

As education is moving toward developing assessments of preparedness for independent learning and 21st-century skills, assessing students' non-cognitive abilities (e.g., attitudes towards feedback-seeking or choosing to revise one's work) is equally important as assessing their cognitive (e.g., learning) abilities across age (e.g., in school, job interviews, etc.). There is an acute need to evaluate how prepared members of society are for independent and lifelong learning, because the ability to adapt to a changing environment by learning from feedback is arguably one of the most important traits of an independent learner. This ability is especially important in educational settings, where students interact with an abundance of feedback from their teachers, peers, or their own work and reflection activities. However, students' attitudes toward feedback valence (i.e., positive or negative) may alter the way they choose to use feedback to enhance their learning outcomes, as it is believed that individuals have a propensity to discount negative information and to readily accept positive

information into their beliefs (Moutsiana et al., 2013). In an empirical research study, this bias against negative information was found to be more pronounced with younger ages, while individuals' engagement with positive information was stable across age (Moutsiana et al., 2013). Although prior research has found age differences in critical feedback processing, it has focused on the valence of feedback rather than on the choice between confirmatory and critical feedback and it has only sampled participants ranging from 9 to 26 years of age (Moutsiana et al., 2013). Other studies showed that younger students learn more from negative than from positive feedback (van der Schaaf, Warmerdam, Crone, & Cools, 2011) but they integrate positive feedback more readily into their beliefs and, thus, alter their behaviors (Crone et al., 2004, 2008; Huizinga, Dolan, & Van der Molen, 2006; van Duijvenvoorde, Zanolie, Rombouts, Raijmakers, & Crone, 2008).

Gaps in the Literature

The current research has identified several theoretical and practical gaps in the related literature. First, there is a current paucity of assessments that can be administered across development. Second, even among the few assessments that can be administered across development, there is a shortage of assessments that work similarly across different ages. Third, although there are many assessments that measure *low-level cognitive constructs* (e.g., IQ, memory, attention, etc.), there are very few assessments that measure *high-level non-cognitive constructs* (e.g., perseverance, mindfulness etc.). Specifically, there is a lack of assessments measuring learning strategies (e.g., individuals' decisions to seek critical feedback and revise their work) that are measurement age-invariant. Fourth, there is a dearth of assessments that examine the *relation* between cognitive and non-cognitive constructs across age. In sum, there is a lack of assessments measuring learning strategies that (1) are measurement age-invariant and (2) have a measurement-age invariant relationship with learning. For instance, no studies have examined the age-invariance of the relation between

feedback-seeking and learning outcomes (i.e., performance and learning). The next subsections offer a more in-depth examination of the gaps identified in this research study. Finally, although much research has focused on feedback that is assigned to students, it is not known how feedback-seeking affects students' learning outcomes (i.e., performance and learning), especially across age. Students need to learn from feedback highlighting performance aspects that could be improved and also recognize situations in which they need to seek feedback proactively to improve their learning. Thus, the goal of the present study is to address the gaps identified in the relevant literature by mainly comparing age-related changes in learning behaviors (e.g., willingness to seek critical feedback) and in the relation between learning behaviors and learning outcomes during childhood, adolescence, and adulthood. A secondary goal of this study is to also examine another learning behavior (i.e., students' choice to revise their work) in relation to age and learning outcomes.

The Research Questions

First, the study focuses on the impact of age on students' learning behaviors (i.e., critical feedback-seeking choice) and learning outcomes (i.e., learning of graphic design principles), respectively. The study investigates whether age affects students' willingness to choose critical feedback when they are presented with critical and confirmatory feedback options about their poster performance. The study also explores whether age affects students' performance on digital posters in the Posterlet game and their appraisal of graphic design principles used correctly or incorrectly on poster exemplars. As a first sub-question, this study aims to discover whether critical feedback-seeking, an elusive non-cognitive construct that is so dependent on individual attitudes, can display an age-invariance property (i.e., whether all test-takers have the same probability of seeking critical feedback, regardless of their age). As a second sub-question, this study investigates whether age is associated with students' learning outcomes (i.e., poster performance and learning of graphic design

principles, respectively). Second, the present study aims to investigate the impact of students' learning behaviors (i.e., the *choice to seek critical feedback* and the *choice to revise* one's work) on learning outcomes. Third, the study investigates whether learning choices and learning outcomes, respectively, vary systematically with age. Fourth, the study explores whether the *relation* between learning choices (i.e., critical feedback-seeking and choosing to revise one's work) and learning outcomes varies systematically with the age of the participants. That is, given a certain level of critical feedback chosen, would all test takers have the same probability of performing and learning well regardless of age or would learning depend on age-group membership?

A large empirical study employing six datasets that ranged from middle-school to college students was conducted to examine the effects of age on students' learning choices, learning outcomes, and on the relation between learning choices and outcomes. In sum, the study poses the following research questions:

1. Does age correlate with learning choices and learning outcomes?
2. Do learning choices correlate with learning outcomes?
3. Do learning choices and learning outcomes, respectively, vary systematically with age?
4. Does the relation between learning choices and learning outcomes vary systematically with age?

The rest of the manuscript is organized as follows. First, the theoretical framework underpinning this research study is presented, together with a review of the relevant literature. Then, a large empirical study that sampled participants starting with age 11 is presented, followed by the results and a discussion of the findings, limitations, and future research directions. Finally, the manuscript concludes with theoretical and practical implications of this research study.

Theoretical Framework

This research draws from the theoretical frameworks of measurement invariance and choice-based assessments to explicate the underlying model of a behavioral game assessment, Posterlet, which we hypothesized to be age-invariant.

Measurement Invariance

Researchers have investigated measurement invariance by measuring psychological constructs across age groups (Alloway, 2012; Dunn & Dunn, 2007; Mungas et al., 2013; Roid & Koch, 2017; Weintraub et al., 2013). Measurement invariance (MI), also known as measurement equivalence, is a statistical property of measurement that indicates whether a given test measures the same underlying construct across groups that also has the same conceptual interpretation across groups. It is a statement that the distribution of observed variables, given the underlying factor scores, is the same in all groups (Lubke, Dolan, Kelderman, & Mellenbergh, 2003). Investigating MI across groups is crucial to any meaningful group comparison and data interpretation based on observed test scores.

Although some tests can be used from the age of 2 through adulthood such as the Woodcock-Johnson Tests of Cognitive Abilities (WJ-COG; Woodcock & Johnson, 1977, 1989; Woodcock, McGrew, & Mather, 2001), most tests cannot be administered across development. For instance, several tests are restricted to late adolescence and beyond, such as the Wechsler Adult Intelligence Scale-III (WAIS-III) used to test the intelligence quotient (IQ) of individuals ranging between 16-89+ years (Bowden et al., 2006). Although many popular psychological assessments have expanded their target administration age range from early childhood to late adulthood, most of these assessment efforts have focused on measuring *low-level cognitive constructs* of foundational abilities, such as IQ (Bowden et al., 2006), memory (Amer, Giovanello, Grady, & Hasher, 2018), and attention (Wirth & Kunzmann, 2018) instead of *high-level non-cognitive constructs* (e.g., feedback-seeking,

revising, etc.). There are very few assessments that measure high-level non-cognitive constructs, such as mindfulness (Droutman et al., 2018), grit (Duckworth & Duckworth, 2016), emotion (Steenhaut, Demeyer, De Raedt, & Rossi, 2018), and personality (e.g., the Personality Assessment Inventory-Borderline Features scale - PAI-BOR – and the Eysenck Personality Profiler – EPP). Recently, research on personality constructs, both normative such as the Big Five Inventory-2 (Shchebetenko, Kalugin, Mishkevich, Soto, & John, 2020) and pathological such as the UPPS-P Impulsive Behavior Scale, has shown these instruments to be age-invariant across the adult lifespan (Argyriou, Um, Wu, & Cyders, 2020). However, a recent systematic review found that it was challenging to establish measurement invariance at certain phases of the lifespan, as many items are age-biased and rely on life experience, which younger participants do not possess (Dong & Dumas, 2020).

Even when they can be administered across ages, few assessments are measurement age-invariant (i.e., yield consistent results across development). For instance, many research studies show an increase in construct scores over time, but others suggest that certain constructs are age-invariant, with corresponding assessments that seem to measure stable traits across age groups.

Finally, relevant to the present research, even fewer studies explored the age-invariance of the relation between non-cognitive constructs and learning. So far, no studies have examined the age-invariance of the relation between feedback-seeking and learning outcomes.

Computer-based Assessment Instrument: Posterlet

The present research study compiled a set of factors to facilitate the positioning of our proposed model within the different classes of age-invariant measurement. Table 1 shows how Posterlet (Cutumisu, Blair, Chin, & Schwartz, 2015), the behavior assessment employed in the current research, is positioned within the measurement invariance literature.

Table 1

The Positioning of the Current Model Within the Different Classes of Age-Invariant Measurements

Criterion	Posterlet
uses the same test for feedback-seeking across age	Yes
obtains the same test results for feedback-seeking across age	Yes
examines the relation across age between feedback-seeking and cognitive* factors (e.g., performance, learning)	Yes
examines the relation across age between feedback and other non-cognitive factors (e.g., revising)	Yes
measures non-cognitive factors (e.g., feedback, perseverance)	Yes
measures cognitive factors (e.g., learning, performance)	Yes
focuses on maturation of cognitive/non-cognitive factors	No
focuses on learning	Yes
focuses on high-level constructs (e.g., feedback, revising, perseverance)	Yes
focuses on low-level constructs (e.g., IQ, memory, attention, etc.)	No

* Here, cognitive refers to high-level constructs, such as learning and performance, rather than memory or IQ.

Figure 1 provides more details about the Posterlet game, including the steps involved in choosing a design theme, designing a poster, choosing three pieces of either positive or negative feedback from virtual characters about that poster, choosing to revise the poster or not, and viewing the number of tickets sold at that poster booth.
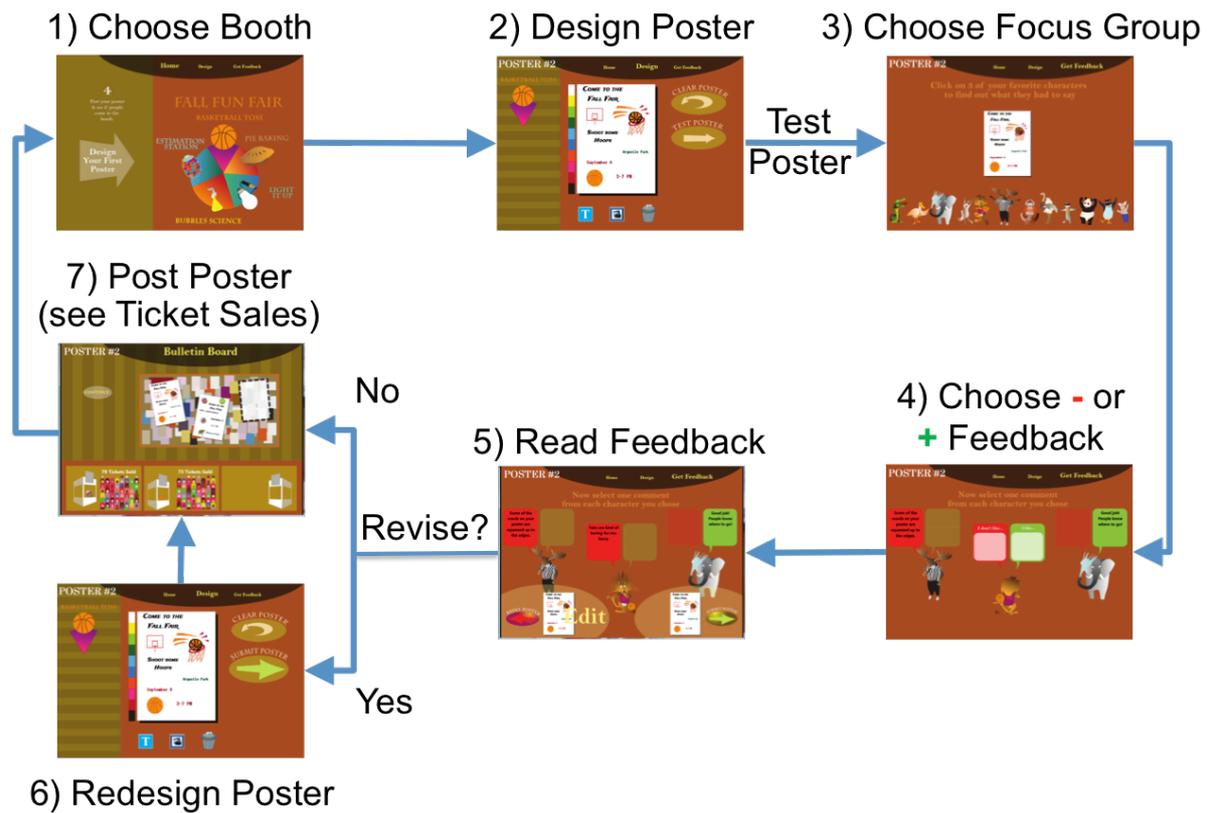
Figure 1. The steps involved in playing the Posterlet game. (Reprinted from Cutumisu et al., 2015.)

This research draws from a game-based assessment approach focusing on the measurement construct of *choice*. Here, choice is defined as an opportunity to decide what and when to learn. The theoretical framework underlying this research is centered on constructivist assessments (Schwartz, Lindgren, & Lewis, 2009), specifically *choice-based* assessments (Schwartz & Arena, 2013), which measure not only students' knowledge but also their choices they exercise regarding their learning. The current study focuses specifically on students' choices and abilities to learn from critical feedback that can help prepare students to become independent learners. It employs the Posterlet game environment as an instance of a *preparation for future learning* (Bransford & Schwartz, 1999; Schwartz & Bransford, 1998) choice-based assessment designed to measure two behaviors important for

autonomous learning: students' choices of seeking *critical feedback* and of *revising* their posters. Concomitantly, it evaluates students' performance outcomes in the assessment environment (e.g., number of tickets sold at the booth in Posterlet, which represents the overall quality of a poster).

The Posterlet game offers choices, opportunities to learn, and an enjoyable environment in which students are more likely to express their authentic learning choices. Previous research found that seeking critical feedback was positively associated with standardized achievement scores for middle-school students, high-school students, and college students (Cutumisu, Blair, Chin, & Schwartz, 2015; Cutumisu, Chin, & Schwartz, 2019; Cutumisu & Schwartz, 2018). Based on these results, the present research hypothesizes that the relation between seeking critical feedback and learning outcomes is positive and it does not vary by age. Thus, the relation between learning choices and learning outcomes is further explored, with a focus on investigating whether the Posterlet assessment is effective for students exceeding 11 years of age. This age cut-off (11 years) was selected because findings from neuroscience research showed that, from around the age of 11 to 13 years, individuals start showing the same decision-making mechanism regarding critical feedback choices as the adults (Peters, Braams, Raijmakers, Koolschijn, & Crone, 2014). More research is needed to discover in which conditions feedback seeking is stable below puberty. Although there is evidence in the literature that the influence of critical feedback on performance differs with age, with 8- to 9-year-old children performing more inaccurately after receiving critical feedback than positive feedback, an fMRI study found that there was no feedback sensitivity difference for 11- to 13-year-olds, and adults performed better after negative than after positive feedback (van Duijvenvoorde et al., 2008). Zhuang, Feng, & Liao (2017) also found that negative feedback led to decreased learning compared to positive feedback and this effect was greater in early adolescents (mean age 12.63 years) than in children (mean age

8.98 years) and adults (mean age 19.79). Other researchers chose the same cut off when they validated the Adolescent and Adult Mindfulness Scale (AAMS) on a comprehensive range of ages starting from 11-year-old children (Droutman et al., 2018). However, in contrast to these studies, the current research examines the situation in which individuals choose the valence of the feedback they receive: confirmatory (i.e., positive) or critical (i.e., negative). Although it is important to monitor students' learning behaviors, research on students' willingness to seek critical feedback is limited and there are no large-scale longitudinal studies on this topic.

Feedback-Seeking Behavior Models

Feedback is largely believed to aid individuals' learning and performance outcomes, because it has the potential to provide new knowledge that corrects errors or gaps in one's understanding (Hattie & Timperley, 2007) and it is most often assigned by experts who know how and when to share the feedback information with the learners. However, the role of feedback-seeking in influencing learning outcomes has been less scrutinized and it has yielded mixed results. A meta-analysis found only a small association between feedback-seeking and performance (Anseel, Beatty, Shen, Lievens, & Sackett, 2015). For instance, a research study found that feedback-seeking behaviors from instructors or peers did not influence students' performance on multiple-choice tests in a blended environment (Hwang & Arbaugh, 2009). They found that competitiveness rather than cooperation seemed to drive feedback-seeking behavior. Although there are many factors that are believed to influence feedback-seeking and its effect on learning and performance outcomes, there is scarce evidence to support these claims in the feedback literature (Anseel et al., 2015).

Ashford and Cummings (1983) introduced a theoretical model of individual feedback-seeking behavior in organizations, drawing from communication processes. They conceptualized feedback as a valuable source of information for individuals, viewing the feedback-seeking process as being driven by motivation and strategies of obtaining valuable

feedback. This model outlines two main feedback-seeking strategies that individuals may choose based on their perceived value of feedback: monitoring (i.e., finding and interpreting feedback cues in the environment) and inquiry (i.e., making direct feedback-seeking requests in the environment). Using these strategies, individuals may ascertain their current behaviors and control their future behavior. For instance, feedback could be used to correct errors and improve performance. Ashford, Blatt, and VandeWalle (2003) outlined three main motives that underlie feedback-seeking behavior: the instrumental motive to perform well or to achieve a goal, the ego-based motive to protect or enhance one's ego, and the image-based motive to protect or enhance others' impressions of oneself. They also emphasized the role of context (e.g., uncertainty, novelty, and change) in influencing these motives. In the context of education, Boud and Molloy (2013) proposed a model of feedback for learning, also based on communication processes, which focused on the role of the learner as a driver for learning through generating and proactively seeking feedback. More recently, also in the context of education, Yan (2020) has emphasized self-assessment as a process that activates students' feedback-seeking behaviors. As in the model proposed by Ashford and Cummings (1983), in this model, learners collect feedback information from internal (through gathering internal-state information from the self; Nicol, 2020) and external (through monitoring and inquiry processes from others in the learning environment) sources.

There are many ways in which feedback-seeking may influence learning outcomes. For instance, protecting one's ego or self-esteem is believed to be an important motivation to seek feedback (Ashford & Cummings, 1983), being achieved through cognitive (e.g., selective memories) and behavioral (e.g., avoiding feedback) defense mechanisms. As such, individuals tend to seek feedback that is congruent with their self-concept (Swann & Read, 1980). Moreover, it was found that individuals who focus on skill development and mastery by persisting through adverse situations are more likely to seek feedback about their

performance, whereas those who seek outside favorable judgements to validate their performance or that prefer to avoid negative judgements, or to show others a lack of competence, rely on maladaptive learning strategies, such as feedback avoidance (Dweck & Leggett, 1988). Only more recently, researchers have started to examine the valence of feedback-seeking (Anseel et al., 2015). There is evidence that individuals tend to seek confirmatory (i.e., positive) feedback more than they seek critical (i.e., negative) feedback (Swann & Read, 1980). This may affect individuals' learning outcomes, as arguably one can learn more from the information associated with criticism than from praise. However, a meta-analysis conducted in an organizational research context found no support for an association between self-esteem and feedback-seeking (Anseel et al., 2015).

Also, the behavioral defense motivation tends to be more salient when an individual expects to perform poorly (Anseel et al., 2015; Ashford & Cummings, 1983). Individuals who expect to perform poorly (i.e., who would require informative feedback the most) tend to avoid diagnostic information (i.e., information about their ability), whereas those who expect to perform well tend to engage in more feedback-seeking behavior (Zuckerman, Brown, Fox, Lathin, & Minasian, 1979). Thus, although individuals may perceive feedback as being useful in correcting errors, they may not always seek it, as they may be motivated to protect their self-esteem. This could also negatively affect individuals' learning outcomes.

Literature Review

Age, Learning Outcomes, and Learning Behaviors

Results of behavioral, psychophysiological, and neuropsychological studies highlighted large developmental differences in learning when individuals learn from positive and negative feedback, respectively. For instance, in a research study, the proportion of hypothesis-testing children increased with age (Schmittmann, van der Maas, Han, & Raijmakers, 2012). A meta-analysis found that individuals in later adulthood show a

significant information processing bias toward positive versus negative information, whereas younger adults show the opposite pattern (Reed, Chan, & Mikels, 2014). It was found that children tend to learn more from positive than from negative feedback, adults tend to learn more from negative than from positive feedback, whereas adolescents tend to show a mixed pattern of learning from feedback that combines those of children and adults (van Duijvenvoorde et al., 2008). This discrepancy across development may be due to differential cognitive control dependent on prefrontal cortex maturation (Crone et al., 2004) as well as to executive function components developing at different rates (Huizinga et al., 2006).

Beyond learning performance considerations, there is some indication in the research literature that individuals' *attitudes* towards learning or learning strategies also tend to vary with age. For example, it was found that grit increases with age (Duckworth & Duckworth, 2016). Also, adolescents and adults were found to be more analytical and efficient learners than young children, employing more learning strategies and metacognitive skills (Pfenninger, 2017). A meta-analysis found that age-related differences in decision-making vary as function of context, such as the learning requirements of the task (Mata, Josef, Samanez-Larkin, & Hertwig, 2011). For instance, a research experiment found that older adults in a positive-mood condition selected more weaknesses to view and spent more time viewing those weaknesses than older adults in the negative-mood condition, while there were no differences across conditions in behavioral results for young adults (Growney & Hess, 2017). An fMRI study examined the influence of age and strategy on neural activity for feedback learning in children and adolescents (i.e., participants between 8 and 25 years old). The findings distinguished four strategy groups that differed in executive functioning efficiency across age groups and found that age differences were important to neural activation beyond strategy (Peters et al., 2014).

An important question for educators is whether individuals' ability to seek feedback and the effectiveness of feedback-seeking on learning performance depends on age. This is an important self-regulated learning (SRL) question, as students' metacognitive ability to decide when and what kind of learning strategies to use is a step towards being an independent, lifelong learner, and it is a predictor of self-determining success in school and life (Schwartz & Arena, 2013). For instance, providing individuals with choice over their feedback can improve learning of motor skills in older adults (Chiviacowsky & Thofehrn, 2017).

Finally, students' feedback uptake is influenced by several factors, including the content of feedback. Specifically, studies found an association between feedback containing concrete suggestions of improving one's work and feedback uptake (van der Pol, van den Berg, Admiraal, & Simons, 2008). Also, feedback targeting the self-regulation and process levels is associated with more feedback uptake (Hattie & Timperley, 2007). Lastly, teachers may model feedback uptake practices to their students to help develop their feedback literacy (Carless & Boud, 2018; Dawson, Carless, & Lee, 2021).

Age-Invariant Test Administration

Some studies suggest that certain tests can be administered to measure *low-level cognitive constructs* across age. For instance, the Woodcock-Johnson Tests of Cognitive Abilities (WJ-COG; Woodcock & Johnson, 1977, 1989; Woodcock et al., 2001) can be used from the age of 2 through adulthood. The Leiter International Performance Scale Third Edition (Leiter-3; Roid & Koch, 2017) is a nonverbal intelligence and cognitive, attentional, and neuropsychological abilities test, updated and re-designed for expanded age ranges (3-75+ years) and specific disabilities. It requires 20-45 minutes to administer and it generates an IQ score. While there are slightly different starting points on each scale given the participants' age, all the same scales are used across all ages. The NIH Toolbox tests, such as

the Peabody Picture Vocabulary Test for language, 4th Edition; PPVT-4 Peabody Picture Vocabulary Test 4th Edition or PPVT-IV (Dunn & Dunn, 2007), the Dimensional Change Card Sort (DCCS) test for executive function together with the Flanker test for attention (Zelazo et al., 2013) can be administered quickly across a wide age range (3–85+ years), allowing for the use of the same metric across all ages. The PPVT-IV is a test of receptive vocabulary and is often used as a proxy for full scale IQ or general developmental level. Finally, the Alloway Working Memory Assessment (AWMA-2) is a fully automated online assessment of working memory (Alloway, 2012) standardized for use with a wide age range (5-69 years). It assesses verbal (auditory) and visuo-spatial short-term memory and working memory.

Although several tests can be administered across age to measure cognitive constructs, there is a paucity of assessments that can be administered to a wide age range to measure *high-level non-cognitive constructs* (e.g., feedback-seeking or revising). For instance, the State-Trait Anger Expression Inventory (STAXI; Spielberger, 1988) measuring trait anger is used for individuals under 30 to over 70 years (Zimprich & Mascherek, 2012). Notably, the Lifespan Self-Esteem Scale (LSE) is another assessment that can be administered across age groups (Harris, Donnellan, & Trzesniewski, 2018). To our knowledge, the Adolescent and Adult Mindfulness Scale (AAMS) is the only scale available that is validated on a comprehensive range of ages from 11-year old children to adults (Droutman et al., 2018). The AAMS scale measures several key components of mindfulness: (1) focus on the present moment, represented by paying attention to surroundings, thoughts, feelings and emotions, (2) being non-reactive, (3) being non-judgmental, and (4) being self-accepting. The operational definition of mindfulness is a process of bringing awareness to moment-to-moment experience and accept any resulting thoughts, feelings, and sensations as they are

(Bishop et al., 2004; Kabat-Zinn, 1990). The language used in the scale is developmentally appropriate across different age groups.

Age-Invariant Tests

In general, there are more tests focusing on the measurement of lower-level cognitive constructs than of higher-level non-cognitive constructs.

In terms of *cognitive constructs*, implicit learning such as grammar is thought to be dependent on early developing structures (Reber, 1993) and acquired in early infancy (Quinn, Westerlund, & Nelson, 2006). Although the WAIS-III was found to be a measure of stable intelligence traits across a wide age range (Bowden et al., 2006), IQ test scores were found to increase over time (Flynn, 1987, 1999). Also, multidirectional age differences were detected in attentional deployment process and outcome (Wirth & Kunzmann, 2018). The NIH Toolbox PPVT-IV vocabulary test (Dunn & Dunn, 2007) showed gradual, linear improvement with age until the mid 50s and then it stabilized (Weintraub et al., 2013). Also, the NIH Toolbox measures of executive function (Dimensional Change Card Sort) and attention (Flanker) for the entire sample and for younger (3–6 years) and older (8–15 years) children yielded moderate to strong positive correlations with age. For many of the assessments that can be administered widely, measurement invariance has not been demonstrated yet. For instance, the Leiter-3 test (Roid & Koch, 2017) can be administered to a wide range of participants, as shown in the previous section but, to our knowledge, there is no evidence that measurement invariance was demonstrated for this assessment. Although the AWMA-2 test (Alloway, 2012) can be used with a wide age range, it is not known empirically whether it satisfies the exigencies of measurement age-invariance.

In terms of *non-cognitive constructs*, personality (e.g., the Personality Assessment Inventory-Borderline Features scale - PAI-BOR and the Eysenck Personality Profiler – EPP) was found to be invariant with age and gender (De Moor, Distel, Trull, & Boomsma, 2009;

Eysenck, Barrett, Wilson, & Jackson, 1992; Picconi, Jackson, Balsamo, Tommasi, & Saggino, 2018). Specifically, given a certain level of the Borderline Personality Disorder (BPD) features, all individuals have the same probability of a certain response on a certain item, regardless of their age or gender. However, grit was found to increase with age (Duckworth & Duckworth, 2016). Although the LSE test measuring self-esteem can be administered across age groups, it is not age-invariant (Harris et al., 2018). Also, mindfulness does not seem to be age-invariant, as there was a small increase in mindfulness with age for university students, while there was no relationship of AAMS with age in a sample of early and mid-adolescents, ages 11 to 18, nor in a sample of early-adolescent middle-school students (ages 11 to 14). Finally, researchers found age-related differences in emotion reactivity (Steenhaut et al., 2018).

Method

We have conducted multiple studies using Posterlet and, overall, we have found consistent results across the various data sets. Moreover, the studies used many different populations, of both high and low socioeconomic status, geographical areas, and age groups. Particularly, the striking consistency of the results across age groups led to the question of whether Posterlet is indeed an age-invariant instrument and feedback-seeking is an age-invariant construct. To find out, we combined the prior data sets and collected additional data to test the effect of age on students' willingness to choose critical feedback and on their learning performance.

Participants and Procedures

This study sampled n = 764 US students from four public middle schools, a high school, and a community college, located in the states of California, Illinois, and New York. In this secondary-data analysis study, we combined data from three public middle schools that were analyzed and reported in our prior work (Chin et al., 2019; Cutumisu, Blair, Chin,

& Schwartz, 2015, 2017) with new data collected from a public middle school, a high school, and a community college. Table 2 shows the school and participant information for the students included in the study. The study was approved by Stanford University's Institutional Review Board, approval number 19824, for the *Choice Activities* project.

Table 2

School and Participant Information

School	Grade (US)	Age	Played Game N	Posttest N	Gender F	M
Central	6	11-12	22	22	13	9
Chicago[*]	6-8	11-14	203	194	43	30
NYC[*]	6-9	11-15	278	231	44	97
Chi/NY[*]	--	--	36	29	--	--
Hillview[*]	7-8	12-14	66	63	35	31
Sequoia	10	15-16	50	45	18	30
Foothill	13-18	15-52	109	103	63	45
Total	**6-18**	**11-52**	**764**	**687**	**216**	**242**

[*] We published results stemming from the data samples marked with an asterisk in Chin et al., 2019; Cutumisu, Blair, Chin, & Schwartz, 2015, 2017. However, this study is fundamentally different from our prior research, as age-invariance is examined for the first time across these samples.

Middle-school and high-school students provided parental consent and then they completed a paper-based assent form. College students completed an online consent form. Then, students played Posterlet (M = 14.45 minutes, SD = 8.51 minutes, n = 727) individually, followed by an online posttest of graphic design principles (M = 3.49 minutes, SD = 2.47 minutes, n = 653). The assessment was carried out in the classroom for middle-school and high-school students and at a location of their choice for college students. Not all students completed the posttest. Therefore, some analyses could not use the full sample. Similarly, no grade level information was available for 36 middle-school students from NYC and Chicago, thus, they were excluded from grade-level analyses. Additionally, only n = 29 (21 females and 8 males) students, n = 18 in Grade 7 and n = 11 in Grade 8, provided consent

from the Hillview school shown in Table 2 out of the total of 66 students. The final sample included n = 727 students (202 females, 219 males, and 306 not declared).

Measurement Instrument: The Posterlet Assessment Game

Posterlet is a game-based behavioral assessment in which players design three digital posters (Cutumisu et al., 2015). It offers students the opportunity to learn graphic design principles through designing posters, seeking feedback, and revising their posters. It is important to note that the Posterlet game was primarily intended to be used as an assessment of learners' willingness to choose critical (i.e., disconfirmatory) feedback when presented with a choice of confirmatory of critical feedback, and not as an instructional game that teaches learners about graphic design principles. Thus, Posterlet was not embedded into the curriculum. Instead, it was administered as a separate task and it was followed by an independent learning test, so that learners' choices could be related to individuals' poster performance and learning of graphic design principles.

Specifically, after designing each poster, players are offered a choice between confirmatory and critical feedback regarding their poster from three interactive game characters selected from a focus group charged with adjudicating the poster, as illustrated in Figure 2.

Figure 2. After submitting each poster, players can choose either confirmatory (i.e., the "I like ..." green box) or critical (i.e., the "I don't like ..." red box) feedback to reveal what each of the three characters thought about the poster.

Overall, students create three digital posters and, thus, have nine choices to seek critical feedback (three choices on each of the three posters). There are 21 graphic design principles that the game's intelligent feedback system uses to evaluate each poster. The feedback system provides feedback (e.g., text is readable, colors make reading difficult, etc.) according to a priority system and it also generates a poster score presented to the player as the number of tickets sold by that poster at the school fun fair, so that players have a general sense of their poster's quality. For instance, if a player used small-size text on a poster, when choosing critical feedback (i.e., by clicking on the "I don't like" red box above one of the three animal characters), the following feedback message may be revealed to the player: "People need to be able to read it. Some of your words are too small." Conversely, if a player uses an appropriate font size and chooses to hear something positive about the poster (i.e., by clicking on the "I like" green box above one of the three animal characters), the following feedback

message may be revealed: "Your poster has big letters. Really easy to read." It is important to note that confirmatory (e.g., "Your poster has big letters. Really easy to read.") and critical feedback (e.g., "People need to be able to read it. Some of your words are too small.") carry equivalent information. Also, once a player receives feedback about one of the 21 graphic design principles on any poster, feedback on that feature (either critical or confirmatory) will not be repeated for the rest of the game.

Finally, after reading the three pieces of feedback, players have one more decision to make: they can choose to revise their poster or submit it as it is. Across the entire game, students have three choices to revise their posters (one choice on each of the three posters).

Measures

Three major classes of continuous measures were employed in this study: (1) learning choices: frequency of choosing critical feedback and frequency of choosing to revise; (2) learning outcomes: poster performance and learning of the graphic design principles; and (3) demographic information: group grade, which constitutes a substitute for age.

Learning Choices

Critical Feedback measures the number of times students choose critical feedback (i.e., by clicking on the "I don't like…" box in the game shown in Step 4 of Figure 1 and in Figure 2) out of a possible maximum of nine (three posters and three opportunities to choose critical feedback per poster). It is important to note that Confirmatory Feedback is a complementary measure, as it represents the maximum number of feedback options across the game (i.e., nine) minus Critical Feedback. Thus, significant positive correlations with Critical Feedback represent significant inverse correlations with Confirmatory Feedback. In prior research, it was found that Critical Feedback was positively associated with other learning behaviors (e.g., the choice to revise one's poster) and learning outcomes (e.g., poster performance and learning of graphic design principles; Chin et al., 2019; Cutumisu, 2019; Cutumisu, Blair,

Chin, & Schwartz, 2015, 2017; Cutumisu, Chin, & Schwartz, 2019; Cutumisu, Schwartz, & Lou, 2020; Cutumisu & Schwartz, 2018).

Revision measures the number of times students choose to revise (i.e., by clicking on the "Edit" arrow in the game shown in Step 5 of Figure 1) out of a possible maximum of three (i.e., three posters and one opportunity to revise per poster).

Learning Outcomes

Poster Quality measures the *in-game performance* on the poster designs. Posterlet evaluated each poster using 21 graphic design principles and produced a poster score per game round (-21 to 21). The Poster Quality measure is the sum of the poster scores (-63 to 63) and is the basis of the number of tickets sold displayed when the play completes each poster (shown in Step 7 of Figure 1).

Posttest measures the *in-game learning* of the design principles from feedback. Students were shown a model poster and they answered two questions by selecting items from a checklist of design principles about what was good and bad, respectively, about that poster. The Posttest measure is the sum of the scores for the two multiple-choice posttest questions (-10 to 10). The answers were scored by assigning a point for each item checked correctly and by subtracting a point for each item checked incorrectly. The correct answers to the two questions are opposites.

Demographic Information

Grade Group is used as a proxy for age, because the schools provided grade, rather than age, information for most of the datasets. This measure was defined as a number ranging from 1 to 5, as follows: 1 = Grade 6 (ages 11-12), 2 = Grade 7 (ages 12-13), 3 = Grade 8 (ages 13-14), 4 = Grades 9 to 12 (high-school students; ages 14-18), and 5 = Grades 13 to 18 (community college students, ages 15-52) to reflect a grouping by educational stages (middle school to college).

Data Analysis Plan

The strategy for the analyses employed by the current study focused on (1) the relation of age with learning choices and learning outcomes, explored using Spearman correlations; (2) the relation between learning choices and learning outcomes; (3) the systematic variation with age of learning choices and learning outcomes, respectively; and (4) the systematic variation with age of the relation between learning choices and learning outcomes. All statistical analyses were conducted using R version 3.6.3 (R Core Team, 2020) and replicated in SPSS.

Results

Descriptive Statistics

First, descriptive analyses were conducted to describe the sample and test the assumptions for the rest of the analyses. Confidence intervals (95%) around the mean for the main variables included in this study are presented in Figure S1 in the *Supplementary Materials*. The mean learning choices and outcomes per Grade Group, along with the sample size of each Grade Group, are displayed in Table 3.

Table 3

Sample Sizes, Means, and Standard Deviations for Critical Feedback, Revision, Poster Quality, and Posttest

Measure	Grade 6 n=155	Grade 7 n=178	Grade 8 n=124	Grade 9-12 n=127	Grade 13-18 n=107
	M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*
Critical Feedback	3.85(2.43)	3.92(2.42)	3.72(2.46)	3.64(2.19)	4.16(2.37)
Revision	1.08(1.13)	1.40(1.16)	1.14(1.20)	0.80(1.02)	1.02(1.18)
Poster Quality	26.97(12.22)	30.57(12.79)	25.73(15.76)	27.96(14.12)	33.42(11.56)
Posttest	1.57(2.27)	2.26(2.78)	2.30(2.49)	1.69(2.67)	2.96(2.42)

The histograms of these variables by grade group are shown in Figures S2 to S5 in the *Supplementary Materials*. In all these figures, the grand sample mean is represented by the

solid gray vertical line, while the dotted black vertical line represents the mean of the grade group (e.g., Grade 6) in each panel. At a glance, this representation of the grand mean and of the individual group mean shows that, for each variable, these means are very similar across age groups. Notably, in the case of the revising behavior, Grade 7 students tend to revise more than the grand mean across age, while high-school students tend to revise less than the grand mean across age.

The histograms of these variables' percentages by grade group are shown in Figures S6 to S9 of the *Supplementary Materials*. These plots show the shape of the distribution for each grade group and variable, providing a percentage measure that is independent of the size of the sample. The box plots of the main variables included in this study per grade group are shown in Figures S10 to S13 (*Supplementary Materials*).

Does age correlate with learning choices and learning outcomes?

The analyses first investigated whether age group predicted learning choices and learning outcomes. Spearman correlations of Grade Group with learning choices (Critical Feedback or Revision) and learning outcomes (Poster Quality or Posttest) were conducted, as these measures were not normally distributed. Table 4 shows the results of all correlations.

Table 4

Correlations of Grade Group with Learning Choices and Learning Outcomes

Measure	Critical Feedback n = 691	Revision n = 691	Poster Quality n = 691	Posttest n = 624
Grade Group	.01	-.09*	.10**	.13**

$^{**}p < .01, ^{*}p < .05$

Results revealed that Grade Group correlated with both learning outcome measures, it inversely and modestly correlated with Revision, but it did not correlate with Critical Feedback. Thus, the older the students are, the more design principles they learn and the

better they perform on their posters, but the less they choose to revise their posters. In sum, while both learning outcomes increased with age and the choice to revise decreased modestly with age, the choice to seek critical feedback was not influenced by age.

Do learning choices correlate with learning outcomes?

Then, analyses investigated whether learning choices were associated with learning outcomes. Spearman correlations of each choice (Critical Feedback and Revision) with each learning outcome (Poster Quality and Posttest) were conducted across the entire dataset. Correlation analyses between the two learning choices (critical feedback-seeking and revision) and between the two learning outcomes (poster performance and posttest), respectively, were also conducted. The correlation coefficients of learning choices with learning outcomes are shown in Table 5.

Table 5

Correlations Between Choices and Learning Outcomes for the Entire Dataset

Measure	Revision n = 727	Poster Quality n = 727	Posttest n = 653
Critical Feedback	.47**	.28**	.19**
Revision	--	.33**	.16**
Poster Quality		--	.34**

$^{**}p < .01$

Results revealed that choices correlated strongly with each other and, to a lesser degree, with both learning outcomes. Additionally, learning outcomes correlated moderately with each other. Thus, the choices to seek critical feedback and to revise seem to be effective choices for learning across a diverse and comprehensive dataset. This analysis supports the internal, convergent validity of our learning measures across an extensive sample. Table S1 and Table S2 show the regression analyses results for Critical Feedback and Revision,

respectively, predicting the learning outcomes and the remaining choice (Revision and Critical Feedback, respectively).

Do learning choices and learning outcomes, respectively, vary systematically with age?

An analysis of variance (ANOVA) was conducted to compare choices across grade groups. Results of analyses of variance showed that there were no differences in critical feedback-seeking between grade groups, with the Levene's test of homogeneity of variances indicating equal variances; however, analyses showed a significant difference in all the other measures, as shown in Table 6. The coefficient estimate information is included in Table S3 of the *Supplementary Materials*.

Table 6

Analyses of Variance Results for Critical Feedback, Revision, Poster Quality, and Posttest by Grade Group, Including the Levene's Test of Homogeneity of Variances

Measure	$F_{(4, 686)}$	p	η^2	Levene's Test	p
Critical Feedback	0.84	.50	.00	0.94	.44
Revision	5.52***	<.001	.03	1.74	.14
Poster Quality	6.50**	<.001	.03	2.42	.047[a]
Posttest	5.31*[b]	<.001	.03	1.29	.27

***p < .001; [a]The Fligner-Killeen test indicates equal variances: 8.68, p = .07; [b]df_2 = 619.

First, there were significant differences in the choice to revise between grade groups, with the Levene's test of homogeneity of variances indicating equal variances. Tukey Honestly Significant Difference (HSD) post-hoc analyses revealed that Grade 9-12 high-school students (mean difference = -0.60, 95% CI[-0.96, -0.24], adjusted p < .001) and Grade 13-18 college students (mean difference = -0.39, 95% CI[-0.77, -0.01], adjusted p = .04) chose to revise significantly less than Grade 7 students.

Second, there were significant differences in poster quality across grade groups, with the Levene's test of homogeneity of variances not indicating equal variances. Tukey HSD

post-hoc analyses revealed that college students significantly outperformed all groups except for 7th-grade students on their poster designs: Grade 6 students (mean difference = 6.45, 95% CI[1.87, 11.03], adjusted $p < .01$), Grade 8 students (mean difference = 7.69, 95% CI[2.88, 12.50], adjusted $p < .001$), and Grade 9-12 students (mean difference = 5.46, 95% CI[0.68, 10.24], adjusted $p = .02$). Also, Grade 8 students (mean difference = -4.83, 95% CI[-9.10, -0.57], adjusted $p = .02$) were significantly outperformed by Grade 7 students.

Finally, there were significant differences in the posttest across groups, with Levene's test of homogeneity of variances indicating equal variances. Post-hoc analyses revealed that college students learned significantly more graphic design principles than the 6th-grade students (mean difference = 1.39, 95% CI[0.47, 2.31], adjusted $p < .001$) and the high-school students (mean difference = 1.27, 95% CI[0.31, 2.22], adjusted $p = .003$).

The first column of Table 7 summarizes the η^2 effect sizes for the learning behaviors and learning outcomes per grade group. All the intra-class correlation (ICC) values were less than .05 (Dyer, Hanges, & Hall, 2005). Thus, there was no justification for the development of a multilevel model (Heck, Thomas, & Tabata, 2013). More detail is provided in the *Multilevel Modeling Assessment* subsection.

Table 7

Descriptive Statistics for the Multilevel Analyses, Including η^2, ICC_1 (Percentage of Variance Due to Groups) and Its Confidence Interval CI_1, and ICC_2 (Reliability of Group Differences) and Its Confidence Interval CI_2.

Variable	η^2	ICC_1	ICC_2	CI_1	CI_2
Critical Feedback	0.00	0.00	-0.20	[-0.005, 0.041]	[-2.353, 0.855]
Revision	0.03	0.03	0.82	[0.007, 0.244]	[0.492, 0.978]
Poster Quality	0.03	0.04	0.85	[0.009, 0.276]	[0.568, 0.981]
Posttest	0.03	0.03	0.81	[0.007, 0.256]	[0.471, 0.977]

Note: *ICC* = Intraclass Correlation Coefficient

Unstandardized and Standardized Effect Sizes

We employed the *bootES* (Kirby & Gerlanc, 2013) R package to apply a *bias-corrected-and-accelerated (BCa)* bootstrapped method with 2000 resamples (i.e., random samples of the same size as the original sample, repeatedly drawn with replacement from the original data sample) to find the 95% confidence intervals (CIs) for the mean difference in each of our four measures (two learning behaviors and two learning outcomes) among the five grade groups. The BCa is recommended as the best method to use among for this purpose (Kirby & Gerlanc, 2013). The bootES ("bootstrap Effect Sizes") method finds bootstrap confidence intervals for between-subjects unstandardized (i.e., the effect size is measured in the original units of measurement) and standardized (i.e., the effect size is measured in standard deviation units) effect-size measures. These include effect sizes for mean effects, mean differences, contrasts, correlations, and correlation differences.

Unstandardized Effect Sizes

The increase in Critical Feedback, Revision, Poster Quality, and Posttest was assessed as a linear function of Grade Group. Table 8 contains the 95% CI (confidence interval) for the mean difference in Critical Feedback, Revision, Poster Quality, and Posttest, respectively, among the five grade groups. This table also includes the standard deviation of the resampled means (SE) and the bias (i.e., the difference between the mean of the resamples and the mean of the original sample).

To test for a linear change in the mean Critical Feedback, Revision, Poster Quality, and Posttest, respectively, as students advanced through the grades, the effect and its associated CI was expressed as a slope of the relationship between the outcome variable and the predictor variable. Table 8 shows the observed slope for Critical Feedback, Revision, Poster Quality, and Posttest, respectively, by Grade Group, including the 95 % CI, bias, and SE. The

findings show that Critical Feedback does not vary linearly by Grade Group, in contrast to the rest of the measures.

Table 8

Assessing the Increase in Critical Feedback, Revision, Poster Quality, and Posttest, Respectively, as a Linear Function of Grade Group

Unstandardized Effect Sizes	Mean Difference	SE	Bias	95% CI		Slope	SE	Bias	95% CI	
				LL	UL				LL	UL
Critical Feedback	.44	.23	.002	-.04	.88	.03	.07	.001	-.09	.17
Revision	.52	.90	.001	.31	.73	-.07	.03	.00	-.14	-.01
Poster Quality	6.18	1.17	.02	3.79	8.39	1.03	.33	.01	.36	1.68
Posttest	1.18	.24	.01	0.68	1.64	.22	.07	.001	.08	.35

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Standardized Effect Sizes

We have also reported standardized measures of effect size as an expression of the effect size in standard deviation units. As we showed before, the mean difference in *Critical Feedback* among grade groups in original units of measure (unstandardized effect size) was 0.44, which does not provide a sense of the magnitude of the effect. Thus, we examined the standardized effect size, Cohen' original *d (sigma)*, which indicated that the difference was slightly less than a fifth of a standard deviation; very similar results were obtained using other methods: (1) Cohen's *d'*; (2) Glass's delta; (3) Robust Cohen's *d*; and (4) Pearson's *r* shown in Table 9. Similarly, the standardized effect size, Cohen' original *d (sigma)*, indicated that the mean difference in *Revision* among grade groups was less than half a standard deviation; very similar results were obtained using the other methods, as shown in Table 9.

Table 9

Standardized Effect Sizes: Critical Feedback and Revision

Effect	Critical Feedback			Revision		
	Standardized Effect Size	95% Bootstrap CI		Standardized Effect Size	95% Bootstrap CI	
		LL	UL		LL	UL
Cohen' original d (sigma)	.19	-.01σ	.40σ	.46	.28σ	.65σ
Cohen's d'	.19	-.004σ	.38σ	.45	.26σ	.64σ
Glass's delta	.18	-.01σ	.38σ	.46	.26σ	.65σ
Robust Cohen's d	.16	-.04σ	.36σ	.45	.27σ	.68σ
Pearson's r	.07[a]	-.01	.14	.18[b]	.11	.25

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit. [a]Bias = .001, *SE* = .04; [b]Bias = 0, *SE* = .04

The mean difference in both the *Poster Quality* and the *Posttest* among grade groups was less than half a standard deviation, as shown by the standardized effect size, Cohen' original *d (sigma)*, in Table 10; very similar results were obtained using the other methods included in this table.

Table 10

Standardized Effect Sizes: Poster Quality and Posttest

Effect	Poster Quality			Posttest		
	Standardized Effect Size	95% Bootstrap CI		Standardized Effect Size	95% Bootstrap CI	
		LL	UL		LL	UL
Cohen' original d (sigma)	.47	.29σ	.63σ	.47	.26σ	.65σ
Cohen's d'	.46	.28σ	.64σ	.46	.26σ	.65σ
Glass's delta	.50	.30σ	.71σ	.52	.29σ	.75σ
Robust Cohen's d	.43	.22σ	.62σ	.48	.28σ	.70σ
Pearson's r	.19[a]	.12	.26	.18[a]	.11	.26

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit. [a]Bias = .001, *SE* = .04

In sum, all the effect sizes were less than half a standard deviation, with the effect size of Critical Feedback across the five grade groups being slightly less than a fifth of a standard

deviation: except for Critical Feedback, all effects were significant. These findings bring additional support that Critical Feedback does not vary with age.

Does the relation between learning choices and learning outcomes vary systematically with age?

Analyses investigated the effect of age on the relation between learning choices and learning outcomes, testing our hypothesis that the relation between learning choices and outcomes does not differ with age. A standard linear regression analysis was conducted that regressed Grade Group, choice, and the interaction term Grade Group by choice on learning outcomes. This analysis explores the simultaneous effect of age, choice (i.e., critical feedback and revision, respectively), and age by choice on learning outcomes. Table 11 and Table 12 show the effect size (i.e., the magnitude of the difference between groups; Levine, 2002) or the percentage of the variance in the dependent variable accounted for by the overall model (Adjusted R^2) and the *t*-values of the Grade Group (i.e., a proxy for age), choice, and Grade Group by choice interaction. When values were missing, cases were excluded pairwise. The confidence intervals for the intercept and each choice predicting the rest of the main variables included in this study are shown in Figures S14 to S19 in the *Supplementary Materials*.

Results shown in Table 11 reveal that, although the model composed of Grade Group, Critical Feedback, and their interaction significantly predicts both learning outcomes and the choice to revise, there are no significant interactions between the grade group and critical feedback predicting learning outcomes or the choice to revise. Thus, the important role of critical feedback choices for learning outcomes and for the choice to revise does not differ with age.

Critical Feedback was a significant individual predictor of both learning outcomes (Poster Quality and Posttest) and of the choice to revise (Revision), which confirms the results yielded by the previous correlation analyses, as shown in Table 11. However, Grade

Group only predicted the Posttest, indicating that the older the students are, the more graphic design principles they learn.

Table 11

Moderator Analysis: Critical Feedback, Grade Group, and Their Interaction Predicting Learning Outcomes and Revision

Effect	B	β	SE	95% CI		t	p
				LL	UL		
Poster Quality (n=727)		$F(3, 687) = 19.72, p < .001, R^2 = $ Adjusted $R^2 = \eta^2 = 0.08$					
(Intercept)[a]	20.06	.00	2.12	15.89	24.23	9.45[***]	<.001
Critical Feedback	1.68	.30	0.46	0.77	2.58	3.63[***]	<.001
Grade Group	1.06	.11	0.70	-0.30	2.43	1.53	.13
Critical Feedback by Grade Group	-0.06	-.04	0.15	-0.36	0.24	-0.39	.70
Posttest (n=653)		$F(3, 620) = 10.35, p < .001, R^2 = 0.05,$ Adjusted $R^2 = 0.04$					
(Intercept)[a]	.37	.00	.43	-0.47	1.21	0.86	.39
Critical Feedback	.33	.30	.10	0.14	0.52	3.36[***]	<.001
Grade Group	.37	.20	.14	0.09	0.64	2.64[**]	<.01
Critical Feedback by Grade Group	-.05	-.17	.03	-0.11	0.01	-1.51	.13
Revision (n=727)		$F(3, 687) = 81.70, p < .001, R^2 = $ Adjusted $R^2 = 0.26$					
(Intercept)[a]	.46	.00	.16	0.14	0.77	2.82[**]	<.01
Critical Feedback	.23	.47	.04	0.16	0.30	6.52[***]	<.001
Grade Group	-.10	-.12	.05	-0.21	0.00	-1.92	.05
Critical Feedback by Grade Group	.01	.04	.01	-0.02	0.03	0.46	.65

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Similarly, results shown in Table 12 revealed that, although the prediction model composed of Grade Group, Revision, and their interaction significantly predicts both learning outcomes and the choice to seek critical feedback, there were no significant interactions between the Grade Group and Revision predicting learning outcomes or the critical feedback-

seeking choice. Thus, the important role of revision choices for learning outcomes or for the critical feedback-seeking choice does not differ with age.

Table 12

Moderator Analysis: Revision, Grade Group, and Their Interaction Predicting Learning Outcomes and Critical Feedback

Effect	B	β	SE	95% CI		t	p
				LL	UL		
Poster Quality (n=727)	$F(3, 687) = 35.43, p < .001, R^2 = $ Adjusted $R^2 = \eta^2 = 0.13$						
(Intercept)[a]	21.97	.00	1.53	18.97	24.97	14.38***	<.001
Revision	3.21	.27	0.95	1.36	5.07	3.40***	<.001
Grade Group	0.82	.08	0.47	-0.10	1.75	1.74	.08
Revision by Grade Group	0.34	.09	0.31	-0.26	0.94	1.13	.26
Posttest (n=653)	$F(3, 620) = 9.08, p < .001, R^2 = $ Adjusted $R^2 = 0.04$						
(Intercept)[a]	1.25	.00	.32	0.62	1.89	3.87***	<.001
Revision	0.20	.09	.21	-0.20	0.61	0.98	.33
Grade Group	0.17	.09	.10	-0.03	0.36	1.68	.09
Revision by Grade Group	0.06	.09	.07	-0.07	0.19	0.96	.34
Critical Feedback (n=727)	$F(3, 687) = 79.07, p < .001, R^2 = 0.26,$ Adjusted $R^2 = 0.25$						
(Intercept)[a]	2.39	.00	.25	1.90	2.88	9.60***	<.001
Revision	1.06	.51	.15	0.76	1.36	6.87***	<.001
Grade Group	0.10	.06	.08	-0.05	0.26	1.35	.18
Revision by Grade Group	-0.00	-.00	.05	-0.10	0.10	-0.05	.96

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Revision was a significant individual predictor of poster performance, but not of learning, as shown in Table 12. This suggests that the more the students choose to revise their posters, the better the posters they design; it also resonates with the results of the previous correlation analyses. As before, Revision was a significant positive predictor of Critical

Feedback, supporting the earlier result that students who are more willing to revise their posters also choose critical feedback more often and vice versa. Grade Group was not an individual predictor of the learning outcomes or of the critical feedback-seeking choice, confirming that seeking critical feedback is an age-invariant choice.

Taken together, these results show that both prediction models were significant and that the importance of choices (choosing critical feedback and choosing to revise) for learning is preserved with age (i.e., it does not differ by age).

Multilevel Modeling Assessment

To triangulate our findings, we also investigated whether multilevel modeling should be used to analyze the data across grade groups. The *lmer* multilevel analytic routine of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in the R Language (R Core Team, 2020) was employed to gain an insight into whether age moderated the relation between seeking learning behaviors and learning outcomes (i.e., whether there was a change in this relation across grade groups). This model-fitting function includes both fixed-effect and random-effect terms to determine the restricted maximum likelihood (REML) estimates of the parameters in linear mixed-effects models. In a multilevel model, the intercept is allowed to vary for each grade-group unit. The intraclass correlation (ICC) was computed (Heck, Thomas, & Tabata, 2013), which constitutes the variance between individuals (between-cluster variance) divided by the total variance (between and within individuals). Thus, ICC is the proportion of variance of individual-level outcome explained by the group (cluster) membership. For random effects, we use ICC to assess effect size (Lorah, 2018). ICC represents a measure of the strength of association, as it constitutes a proportion of variance.

The variance between individuals (variance of the intercept) for age moderating the relation between seeking critical feedback and performance was 6.454, while the variance of the residual was 165.578. Thus, ICC = 6.454 / (6.454 + 165.578) = .037. This value does not

exceed the 5% threshold (Dyer et al., 2005), so we cannot assume statistically significant variability in intercepts between students. Therefore, this does not justify the development of a multilevel model (Heck, Thomas, & Tabata, 2013). This analysis indicates that the variation due to age (Grade Group) is small, suggesting that the difference in intercepts is not significant and that the relation between seeking Critical Feedback and Poster Quality is age-invariant. An analysis for the Posttest yielded similar results: ICC = .2017 / (.2017 + 6.2929) = .031, indicating that the relation between seeking critical feedback and learning of graphic design principles (as measured by the Posttest) is also age-invariant.

A similar analysis for revision and performance: ICC = 8.002 / (.8002 + 156.300) = .048 and for revision and the posttest: ICC = .2319 / (.2319 + 6.3292) = .035, respectively, indicated that the relation between choosing to revise and learning outcomes (i.e., poster performance and learning of graphic design principles) is also age-invariant.

For completion, we also conducted this analysis for performance and grade group predicting the posttest: ICC = 0.1409/(0.1409 + 5.8675) = .02 as well as for the critical feedback-seeking and grade group predicting revision: ICC = 0.03689/(0.03689 + 0.96043) = .04.

Taken together, all the results indicate that the creation of a multilevel model is not justified, as all the ICC values were below the 5% threshold, so there is no statistically significant variability in intercepts between students across grade groups for any of the measures included in this study.

Discussion and Future Work

This study investigated students' choices to seek critical feedback and to revise their posters depending on their age. It also aimed to discover whether the hypothesized relation between seeking critical feedback and learning performance varied systematically with age.

Age Versus Learning Choices and Learning Outcomes

Results revealed that the older the students were, the more they learned. Specifically, the older the learners, the better they performed on their posters and the more they learned graphic design principles. As other research has found, it could be that older participants have better mechanisms to monitor the consequences of their actions and control their behaviors to improve their outcomes (Luszcz, 2011). This result aligns with typical developmental changes, but no other studies have explored the relations among the learning choices to *seek feedback* and to revise, learning performance, and learners' age. However, age was found to correlate positively with feedback-learning performance (i.e., a core skill for cognitive flexibility) in a study with students aged 8 to 25 years (Peters, Van der Meulen, Zanolie, & Crone, 2017), but feedback was assigned to the students in that study. In contrast, the present research focuses on feedback that is freely chosen by the students. Additionally, the present study revealed that the older the students, the less they chose to revise their posters.

Most importantly, the present study shows that the choice to seek critical feedback is age-invariant. Specifically, the study showed that (a) age does not correlate with seeking critical feedback and (b) there were no significant differences between grade groups on seeking critical feedback. This means that individuals of different ages, starting with age 11, exhibit the same patterns of critical feedback-seeking (i.e., they choose about the same quantity of critical feedback). While both learning outcomes (performance and learning) significantly increased systematically with age and the choice to revise decreased modestly with age, the choice to seek critical feedback was not influenced by age.

This result suggests that feedback-seeking is stable through development, with both young children and adults using a similar mechanism when choosing between critical and confirmatory feedback. Thus, even if there are individual differences in the sample, with students from different cities and different schools having a wide range of income (e.g.,

individual parental income in the school neighborhood varied from \$57,717 to \$204,250), no differences in choosing critical feedback across age, starting in middle-school, were found.

Importantly, this study shows that the same assessment (i.e., the same version of the Posterlet game) was used across all samples, with consistent results across age groups. To our knowledge, this is the first time that age-invariance is demonstrated for feedback-seeking.

Learning Choices Versus Learning Outcomes

Results show that both learning choices (i.e., seeking critical feedback and revising) correlate with learning outcomes (poster performance, which represents students' learning during the assessment, and posttest, which represents students' learning of the graphic design principles assessed after the Posterlet game) across a diverse and comprehensive dataset. Moreover, both learning outcomes correlate with each other, providing internal, convergent validity for the measures considered in this study.

No other research examined the relation between students' *choices* to seek critical feedback and their *choices* to revise their work. Although results show that choosing to seek critical feedback and to revise strongly correlate with each other, it is not known whether critical feedback caused students to revise their posters or whether those who seek critical feedback are more likely to revise. A future causal study will determine the direction of this relation between learning choices. Meanwhile, this study found that students' learning choices to seek critical feedback and to revise their work were strongly associated with each other.

Choosing critical feedback correlates with both in-game (i.e., poster performance) and out-of-game (i.e., independent posttest) learning measures, possibly due to differential processing between confirmatory and critical feedback, or to other psychological factors. Confirmatory feedback may highlight design principles that students already know and use well, whereas critical feedback may bring forward new knowledge related to graphic design.

More research is necessary to elucidate the relation between learning choices and learning outcomes. Meanwhile, these results are supported by other researchers who found that participants learned more from negative that from positive feedback (van der Schaaf et al., 2011).

Results also show that revising after critical feedback-seeking correlates with both learning measures, though to a weaker degree with posttest learning. Revision improves performance on the poster designs, presumably because by revising, one may fix any potential issues and avoid repeating them. Revision improves, albeit to a lesser extent, learning of the graphic design principles, perhaps because by revising, one may revisit some principles and recognize them readily when they are presented on the posttest. Some of the principles on the posttest were never presented as feedback, as students had limited feedback opportunities and unique poster designs. Thus, revision is an effective learning choice for improving performance, but not necessarily to the same extent for improving learning. Merely performing the behavior of revising does not guarantee learning of graphic design principles, but it may improve performance. A future study will be designed to single out the effect of revision on performance and learning.

The current results support the initial hypothesis that learning choices to seek critical feedback and to revise are helpful choices for learning and performance, regardless of learners' age. One possible explanation for these results is that constructive feedback helps learners improve their performance by filling the gap between what they know and what they need to know, regardless of their age. Recent studies have found that revision is part of the mechanism that links seeking critical feedback and improved performance and learning, as revision fully mediated the link between critical feedback-seeking and learning outcomes (Cutumisu & Lou, 2020; Cutumisu, Schwartz, & Lou, 2020).

Age, Learning Choice, and the Interaction between Age and Learning Choice Versus Learning Outcomes

The findings reveal that the relation between learning choices (i.e., choosing critical feedback and choosing to revise) and learning outcomes does not vary by age. Thus, seeking critical feedback is beneficial for learning (i.e., the more the participants choose critical feedback in Posterlet, the more they learn graphic design principles, as measured by the Posttest and the better posters they design, as measured by the Poster Quality), regardless of age. Moreover, the relation between critical feedback-seeking and learning outcomes is age-invariant, being also stable throughout development, just like the behavior of seeking critical feedback. These results suggest that age does not necessarily constitute an advantage in processing and using productive learning behaviors (e.g., critical feedback-seeking and revising). Although the oldest participants performed better on the posters and learned more than most of the other grade groups, they still learned at the same rate and performed similarly on their posters as the youngest participants when choosing the same amount of critical feedback.

One important implication of this result is that, starting in middle school, individuals who choose a certain quantity of critical feedback have the same probability of learning from that feedback throughout their life. Specifically, there is no significant difference in how older and younger participants learn from critical feedback-seeking, and all learners improve their learning outcomes by seeking critical feedback (i.e., all learners benefit in the same way from seeking critical feedback). Importantly, this result also suggests that young children and adults use a similar mechanism when choosing and using feedback for learning enhancement.

Also, there was no interaction between age and seeking critical feedback in predicting revision. At the same time, there was no interaction between age and revision predicting critical feedback-seeking. Therefore, although students choosing critical feedback are more

likely to revise their posters, age is not moderating this relationship. It is possible that students who seek critical feedback are compelled to revise their work based on that feedback. As students choose critical feedback before they decide whether to revise their posters, it is possible that students' decision to revise may be influenced by the amount of critical feedback chosen. Concomitantly, it is also possible that students who know that revising is an effective learning strategy, being emphasized by teachers or schools (Growney & Hess, 2017), actively seek information that would help them identify the aspects that they need to revise and thus they tend to seek critical feedback more often. Follow-up studies are needed to clarify the role of revision in relation with critical feedback-seeking.

Moreover, results show that students choosing critical feedback are more likely to perform better on the poster task and learn more, but that age does not moderate the relation between seeking critical feedback and the other learning behaviors (revision) or outcomes (poster performance or learning). Similarly, students choosing to revise are more likely to perform better on the posters, but age does not moderate the relation between revising and poster performance. An analysis by level of grade group revealed that Grade 8 students who revised more than one poster also designed significantly better posters than those who revised less. Finally, other research has found that the influence of critical feedback on performance differs with age (Zhuang et al., 2017), with adults performing better after negative than after positive feedback (van Duijvenvoorde et al., 2008). Another study that estimated the learning rate of children, adolescents, and adults found that the learning rate based on negative feedback decreased with age, whereas the learning rate based on positive feedback did not vary with age (van den Bos, Cohen, Kahnt, & Crone, 2012). Two possible explanations for the discrepancy of the findings regarding the role of age in the relationship between negative feedback and learning could stem from the types of learning involved, which may recruit brain regions that differ in their sensitivity to negative feedback, as well as from the

motivational level of the participants (Zhuang et al., 2017). Finally, although revision is not age-invariant, as it tends to decrease with age, the relation between choosing to revise one's poster and learning behaviors (critical feedback-seeking) and outcomes (poster performance and learning of graphic design principles) is age-invariant. Taken together, these results show that the contribution of effective learning choices (e.g., seeking critical feedback and choosing to revise) to learning is age-invariant, suggesting that such learning behaviors are essential for lifelong learning. Overall, this study constitutes a demonstration that both *choosing* critical feedback and *choosing* to revise yield better learning to the same degree for all ages investigated.

Although research shows differential learning from feedback across development (learning from critical feedback increases with age) when students are *assigned* feedback (Peters et al., 2014), the current study did not find age differences in the relation between critical feedback-seeking and learning when students *choose* their feedback. This may be due to the agency that participants exercise when they decide when and what kind of feedback to choose about their work. The findings of the present study have implications for the design of interactive learning environments that include automated feedback, as they pave the way to more personalized learning approaches and the need for these environments is highlighted in much of the recent relevant literature (Crosslin, 2018).

Limitations

This study was correlational and not causal, thus, the modest differences in revision across different ages could be a function of students' different cities, parental income, school curricula, teachers, climate, or other factors. It is also possible that there is a latent variable (e.g., persistence) that drives the correlations between feedback, revision, and learning performance. Moreover, it would have been preferable to include a broader set of measures that would enable the demonstration of divergent validity, because this study measures

something that is not captured by other common assessments. It might be useful to know whether seeking critical feedback shows a different pattern of correlation with learning outcomes than other relevant predictors (e.g., self-efficacy, mindset, etc.). Now that an assessment which measures learning choices was developed and evidence that choices are related to learning in the assessment environment was found, it will be possible for researchers to determine why students of different ages seek critical feedback or revise, and for educators to evaluate whether a curriculum prepares students to make such independent learning choices. However, a recent study comparing the learning choices and outcomes of students who could choose their feedback valence and those who were assigned the same feedback valence, amount, and order, revealed no differences between conditions (Cutumisu & Schwartz, 2018). In another study, it was found that mindset did not influence students' learning choices and learning outcomes, but that it influenced students' responses to their chosen strategies (Cutumisu, 2019). The latter result contrasts the belief that learning goal orientation would promote feedback-seeking (Anseel et al., 2015) through a growth mindset (Dweck & Leggett, 1988). Taken together, these results imply that there are no other factors such as mindset that influence students' learning choices, as students displayed the same learning outcomes in both conditions (i.e., the same amounts of critical feedback-seeking lead to the same learning outcomes, regardless of condition).

In this study, we used five grade groups as a proxy for age. Future studies will collect age information from the participants and explore the relationship between age and learning choices and outcomes using a finer granularity of the age measure.

Some schools did not provide students' academic achievement scores. Thus, it was not possible to explore the correlation of learning choices with both internal (in the assessment environment) and external (outside the assessment environment) learning outcomes. A future study will examine the relations between learning choices and both in-game and in-school

learning outcomes to provide external validation of the assessment. Prior research showed that both critical feedback-seeking and revision correlated with standardized mathematics and reading scores in two states (NY, Illinois) for Grade 6 to 9 students (Cutumisu, Blair, Chin, & Schwartz, 2015). A recent study showed that prior academic achievement was positively related to Grade 6 students' spontaneous use of critical feedback-seeking (Cutumisu, Schwartz, & Lou, 2020). Moreover, critical feedback-seeking was positively associated with students' performance on the poster design task through revising their posters, regardless of their prior academic achievement.

Finally, it was suggested that individuals' acceptance of feedback and response to feedback are different when feedback is sought versus received (Ashford & Cummings, 1983). Indeed, a recent study comparing the learning outcomes of college students who chose versus were assigned the same amount of negative feedback found that, while critical feedback and revision choices were positively associated with performance when students choose their feedback, critical feedback was negatively associated with learning when students were assigned their feedback (Cutumisu & Schwartz, 2018). Future studies will examine the psychological differences in protecting one's ego between making free choices (e.g., *choosing* to seek critical feedback) and being assigned choices (e.g., *receiving* critical feedback without requesting it) as they relate to learning and age differences. Choosing critical feedback may diffuse ego threat, whereas being assigned critical feedback by the game characters may lead to less learning than enabling students to choose critical feedback. This issue is relevant to many instructional technologies.

This research puts forward a pressing question that should be addressed by all assessment efforts. Given a measure of learning, is there also a way to improve outcomes by this measure? This study does not have an answer to the question of the best way to help students learn to choose critical feedback and to revise. However, now that these choices can

be measured, it should be possible to investigate how to inculcate an attitude that embraces critical feedback as a chance to learn rather than a reflection of one's personal worth. Assessing choices provides a new approach for evaluating process skills that are elusive to more traditional testing, but are of great interest to many educators. This research views choice as an important outcome of learning across different ages, instead of as a source of motivation and self-selection.

Educational Implications

Theoretical Implications

The present study makes several important theoretical contributions. First, no other studies have assessed feedback-seeking across development; this study examines feedback-seeking starting in middle school. As interest in feedback interventions continues to grow, so will the importance of developing instruments and of measuring the feedback-seeking construct across populations. This study examines the Posterlet assessment game, a choice-based assessment of learning, as well as students' performance, feedback-seeking behavior, and choices to revise their own work. Posterlet is a measurement-invariant assessment of feedback-seeking choices across age, starting in middle school. The exact same instrument, Posterlet, was designed to be administered to several age groups and it has indeed been administered to 727 students ranging from age 11 to adulthood, measuring the same underlying construct (i.e., the critical feedback-seeking choice) and being interpreted in a similar conceptual manner across different age groups with the same consequences on learning outcomes. In contrast to assessments such as the Leiter-3 test, Posterlet measures a high-level, non-cognitive construct, it requires a shorter administration time, it does not require a different starting point for different age groups, using exactly the same game for all the age groups tested, and it now provides evidence of measurement invariance for the feedback-seeking construct. Thus, similar with the NIH Toolbox PPVT-IV vocabulary test

(Dunn & Dunn, 2007), Posterlet can be administered across a wide age range and, importantly, it enables the use of the same metric for all ages starting with 11-year-old children.

Second, this research shows for the first time that students' willingness to choose feedback is age-invariant. The distribution of choices (critical feedback-seeking and choosing to revise) is also age invariant across a large cross-sectional sample. This shows that the choice to seek critical feedback is stable across age. This result is in contrast to findings in organizational research. For instance, a meta-analysis of 69 articles of 30 years of feedback-seeking research, which excluded publications in an educational context, revealed that age was negatively related to feedback-seeking behavior (Anseel et al., 2015). This discrepancy may stem from the different research contexts, the digital game environment in which feedback is only seen by the player versus a real situation in an organization, the feedback being conveyed by virtual characters versus a person, the task domain, the goal orientation (learning versus performance goal orientation), or the different types of repercussions.

Third, to our knowledge, no other instrument has examined the age invariance of the relation between a strategy and its resulting learning outcomes. Posterlet offered learning opportunities concomitantly with assessing students, so it was possible to measure students' performance outcomes as well as their learning choices while they were designing posters. An important result of this study is also the positive association between critical feedback-seeking and learning outcomes, which yielded mixed results in the feedback-seeking literature (Anseel et al., 2015).

Fourth, this research shows for the first time that the benefits gained by seeking critical feedback for learning are also age-invariant across a large developmental sample, using strong empirical evidence.

Finally, in the light of the theoretical model of individual feedback-seeking behavior put forward by Ashford and Cummings (1983), the result of the current study showing that critical feedback-seeking is positively associated with learning regardless of age can be used to develop individuals' inquiry skills for feedback-seeking at any age. Thus, learning environments could promote feedback-seeking of critical information as a way to remedy gaps in one's understanding, which are inherent during the learning process, and to improve one's learning outcomes, without negatively affecting one's self-esteem. Games such as Posterlet offer constructive feedback via virtual characters, which has the potential to attenuate the stigma of negative feedback.

Practical Implications

In contrast to other assessments reviewed in this study, the Posterlet assessment (1) measures a high-level construct, feedback-seeking, and its relation with learning, exploring the two across age groups; (2) is a digital online game-based assessment that sustains the interest of the players, so it can be administered for a wide range of ages; (3) yields the same results across development starting in middle school, which indicates that a high-level construct such as feedback-seeking has a constant distribution across age; (4) takes about 10-15 minutes to administer; (5) can scale to simultaneously test a large group of students; (6) requires no training to administer, as the game URL can be distributed to participants who are instructed via a short click-through tutorial at the beginning of the game; (7) requires no prior content knowledge (i.e., graphic design), so it has the advantage that it does not have to be embedded into a specific curriculum; and (8) is a creative, open-ended task that affords many different solutions. The game is non-verbal, relying on mostly visual elements and keeping the text at a minimum. The scoring of the performance on the task is fully automated: individuals can view their score in the form of tickets sold at their poster booth.

Professionals may use an age-invariant instrument such as Posterlet to gain a more objective understanding of individuals' learning behaviors (e.g., willingness to choose critical feedback), to prepare for responding to children's various developmental demands, and to analyze the effectiveness of their behavioral interventions. This research has implications for curriculum design that would afford learners opportunities for development of monitoring and inquiry skills to recognize feedback in their environment, to judge its value for a specific learning goal, and to actively seek the feedback valence that leads to better learning outcomes. As the learning context is important in influencing feedback-seeking motives and subsequent consequences, the classroom culture may shift to encouraging feedback-related inquiry. For instance, instructors may create environments abundant in feedback choices to encourage learners to be more proactive about their learning, they may seek feedback from the students more often, and may implement any changes as a result of the feedback to model feedback-seeking behavior, especially as research shows that instructors need to model good feedback uptake behaviors (Carless & Boud, 2018; Winstone & Carless, 2019). Thus, learners may benefit more from the critical feedback they choose, compared to the more traditional approach where critical feedback is assigned to them by an instructor. This is an important distinction, as feedback that is chosen by the learner is more likely to be optimized to each learner's performance level, while feedback that is assigned to the learner may not be aligned with the performance level of the learner. Specifically, learners may more readily accept and heed critical feedback they choose rather than the same amount they are assigned, as they may not necessarily be prepared to receive and use unwanted criticism to improve their work, if they are not ready for it. Instructors may also benefit from accessing a dashboard (e.g., based on the Posterlet assessment) that displays in real time the amount of critical feedback each student chooses. Thus, instructors may work with students who are reluctant to choose critical feedback, find ways to uncover the reasons for this reluctance, and

intervene using design thinking ("fail early, fail often") or growth mindset strategies. Prior research has shown that students can be taught design thinking strategies (choosing critical feedback) and that, importantly, they can transfer these strategies to other learning environments (Chin et al., 2019).

Conclusions

The study presents empirical evidence that choosing critical feedback is both age-invariant and it also predicts better learning and performance regardless of individuals' age, starting in middle school. Specifically, the relation between feedback seeking and learning outcomes is also age-invariant starting in middle school. Thus, although learning outcomes improve with age, the relation between learning choices and outcomes does not vary with age. This is an important result, as it is not common in the current literature for such a high-level construct as feedback seeking to have a constant distribution across ages starting in middle school. Also, age does not moderate the relation between participants' willingness to seek critical feedback and their willingness to revise their posters. The results of this novel examination of critical feedback and revision choices suggest that effective learning choices may be added to the repertoire of independent learners of any ages who will likely make such choices beyond school to improve their learning outcomes.

Acknowledgement

References

Alloway, T. P. (2012). Alloway Working Memory Assessment 2 (AWMA-2).

Amer, T., Giovanello, K. S., Grady, C. L., & Hasher, L. (2018). Age differences in memory for meaningful and arbitrary associations: A memory retrieval account. *Psychology and Aging, 33*(1), 74-81.

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How are we doing after 30 years? A meta-analytic review of the antecedents and outcomes of feedback-seeking behavior. *Journal of Management, 41*(1), 318-348. https://doi.org/10.1177/0149206313484521

Argyriou, E., Um, M., Wu, W., & Cyders, M. A. (2020). Measurement invariance of the UPPS-P impulsive behavior scale across age and sex across the adult life span. *Assessment, 27*(3), 432-453.

Ashford, S. J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management, 29*(6), 773-799. https://doi.org/10.1016/S0149-2063(03)00079-5

Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance, 32*, 370-398. https://doi.org/10.1016/0030-5073(83)90156-3

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https:/doi.org/10.18637/jss.v067.i01

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., . . . Velting, D. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice, 11*(3), 230-241.

van den Bos, W., Cohen, M. X., Kahnt, T., & Crone, E. A. (2012). Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral cortex, 22*(6), 1247-1255. https://doi.org/10.1093/cercor/bhr198

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Bowden, S. C., Weiss, L. G., Holdnack, J. A., & Lloyd, D. (2006). Age-related invariance of abilities measured with the Wechsler Adult Intelligence Scale-III. *Psychological Assessment, 18*(3), 334.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education* (pp. 61-100). Sage Publications Sage CA.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education.* https://doi.org/10.1080/02602938.2018.1463354

Chin, D. B., Blair, K. P., Wolf, R. C., Conlin, L. D., Cutumisu, M., Pfaffman, J., & Schwartz, D. L. (2019). Educating and measuring choice: A test of the transfer of design thinking in problem solving and learning. *Journal of the Learning Sciences (JLS), 3*(28), 337-380. https://doi.org/10.1080/10508406.2019.1570933

Chiviacowsky, S., & Thofehrn, H. (2017). Choices over feedback enhance motor learning in older adults. *Journal of Motor Learning and Development,* 1-24.

Crone, E. A., Ridderinkhof, K. R., Worm, M., Somsen, R. J., & Van Der Molen, M. W. (2004). Switching between spatial stimulus–response mappings: a developmental study of cognitive flexibility. *Developmental Science, 7*(4), 443-455. https://doi.org/10.1111/j.1467-7687.2004.00365.x

Crone, E. A., Zanolie, K., Van Leijenhorst, L., Westenberg, P. M., & Rombouts, S. A. (2008). Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective, & Behavioral Neuroscience, 8*(2), 165-177.

Crosslin, M. (2018). Exploring self-regulated learning choices in a customisable learning pathway MOOC. *Australasian Journal of Educational Technology, 34*(1).

Cutumisu, M. (2019). The association between critical feedback seeking and performance is moderated by growth mindset in a digital assessment game. *Computers in Human Behavior, 93*, 267-278. https://doi.org/10.1016/j.chb.2018.12.026.

Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2015). Posterlet: A game-based assessment of children's choices to seek feedback and to revise. *Journal of Learning Analytics, 2*(1), 49-71.

Cutumisu, M., & Lou, N. M. (2020). The moderating effect of mindset on the relation between university students' critical feedback-seeking and learning. *Computers in Human Behavior, 112*. https://doi.org/10.1016/j.chb.2020.106445, 106445.

Cutumisu, M., & Schwartz, D. L. (2018). The impact of critical feedback choice on students' revision, performance, learning, and memory. *Computers in Human Behavior, 78*, 351-367. https://doi.org/10.1016/j.chb.2017.06.029.

Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2017). Assessing whether students seek constructive criticism: The design of an automated feedback system for a graphic design task. *International Journal of Artificial Intelligence in Education, 27*(3), 419-447. https://doi.org/10.1007/s40593-016-0137-5

Cutumisu, M., Chin, D. B., & Schwartz, D. L. (2019). A digital game-based assessment of middle-school and college students' choices to seek critical feedback and to revise. *British Journal of Educational Technology, 50*(6), 2977-3003. https://doi.org/10.1111/bjet.12796

Cutumisu, M., Schwartz, D. L., & Lou, N. M. (2020). The relation between academic achievement and the spontaneous use of design-thinking strategies. *Computers & Education, 149*. https://doi.org/10.1016/j.compedu.2020.103806, 103806.

Dawson, P., Carless, D., & Lee, P. P. W (2021). Authentic feedback: Supporting learners to engage in disciplinary feedback practices. *Assessment and Evaluation in Higher Education, 46*(2), 286-296. https://doi.org/10.1080/02602938.2020.1769022

De Moor, M. H., Distel, M. A., Trull, T. J., & Boomsma, D. I. (2009). Assessment of borderline personality features in population samples: Is the personality assessment inventory-borderline features scale measurement invariant across sex and age? *Psychological Assessment, 21*(1), 125-130.

Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences, 160*, 1-23. https://doi.org/10.1016/j.paid.2020.109956, 109956.

Droutman, V., Golub, I., Oganesyan, A., & Read, S. (2018). Development and initial validation of the adolescent and adult mindfulness scale (AAMS). *Personality and Individual Differences, 123*, 34-43.

Duckworth, A., & Duckworth, A. (2016). *Grit: The power of passion and perseverance.* Findaway World LLC.

van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E., & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience, 28*(38), 9495-9503. https://doi.org/10.1523/JNEUROSCI.1485-08.2008

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test manual* (4th ed.), Pearson Assessments, Minneapolis, MN.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and

    personality. *Psychological Review, 95*(2), 256–273. https://doi.org/10.1037/0033-

    295X.95.2.256

Dyer, N., Hanges, P., Hall, R. (2005). Applying multilevel confirmatory factor analysis

    techniques to the study of leadership. *The Leadership Quarterly, 16*, 149-167.

    https://doi.org/10.1016/j.leaqua.2004.09.009

Eysenck, H. J., Barrett, P., Wilson, G., & Jackson, C. (1992). Primary trait measurement of

    the 21 components of the PEN system. *European Journal of Psychological Assessment,*

    8(2), 109-117.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure.

    Psychological Bulletin, 101(2), 171-191.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American*

    Psychologist, 54(1), 5-20.

Growney, C. M., & Hess, T. M. (2017). Affective influences on older adults' attention to

    self-relevant negative information. *The Journals of Gerontology: Series B*.

Harris, M. A., Donnellan, M. B., & Trzesniewski, K. H. (2018). The lifespan self-esteem

    scale: Initial validation of a new measure of global self-esteem. *Journal of Personality*

    Assessment, 100(1), 84-95.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*

    77(1), 81–112. doi:10.3102/003465430298487

Heck, R. H., Thomas, S. L., & Tabata, L. (2013). *Multilevel and longitudinal modeling with IBM SPSS.* Routledge.

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Hwang, A., & Arbaugh, J. B. (2009). Seeking feedback in blended learning: Competitive versus cooperative student attitudes and their links to learning outcome. *Journal of Computer Assisted Learning, 25*(3), 280–293. https://doi.org/10.1111/j.1365-2729.2009.00311.x

Kabat-Zinn, J. (1990). Full catastrophe living: Using the wisdom of your mind to face stress, pain, and illness. 58-72. http://lelandshields.com/Meditation/Full%20Catastrophe%20Living%20-%20Kabat-Zinn.pdf

Kirby, K.N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods, 45*, 905–927. https://doi.org/10.3758/s13428-013-0330-5

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612-625.

Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-scale Assessments in Education, 6*(8). https://doi.org/10.1186/s40536-018-0061-2

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within-and between-group differences and measurement invariance in the common factor model. *Intelligence, 31*(6), 543-566.

Luszcz, M. (2011). Executive function and cognitive aging. *Handbook of the psychology of aging* (7th ed.). Elsevier.

Mata, R., Josef, A. K., Samanez-Larkin, G. R., & Hertwig, R. (2011). Age differences in risky choice: A meta-analysis. *Annals of the New York Academy of Sciences, 1235*(1), 18-29.

Moutsiana, C., Garrett, N., Clarke, R. C., Lotto, R. B., Blakemore, S., & Sharot, T. (2013). Human development of the ability to learn from bad news. *Proceedings of the National Academy of Sciences, 110*(41), 16396-16401.

Mungas, D., Widaman, K., Zelazo, P. D., Tulsky, D., Heaton, R. K., Slotkin, J., & Gershon, R. C. (2013). VII. NIH Toolbox Cognition Battery (CB): Factor structure for 3 to 15 year olds. *Monographs of the Society for Research in Child Development, 78*(4), 103-118.

Nicol, D. (2020). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 1-23. https://doi.org/10.1080/02602938.2020.1823314

Peters, S., Braams, B. R., Raijmakers, M. E. J., Koolschijn, P. C., & Crone, E. A. (2014). The neural coding of feedback learning across child and adolescent development. *Journal of Cognitive Neuroscience, 26*(8), 1705-1720.

Peters, S., Van der Meulen, M., Zanolie, K., & Crone, E. A. (2017). Predicting reading and mathematics from neural activity for feedback learning. *Developmental Psychology, 53*(1), 149-159. https://doi.org/10.1037/dev0000234

Pfenninger, S. E. (2017). Not so individual after all: An ecological approach to age as an individual difference variable in a classroom. *Studies in Second Language Learning and Teaching, 7*(1), 19-46.

Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., & Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck personality questionnaire-short form (EPP-S). *Personality and Individual Differences, 123*, 76-80.

van der Pol, J., van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*(4), 1804-1817. https://doi.org/10.1016/j.compedu.2008.06.001

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science, 17*(1), 59-66.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundations for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious oxford psychology series, № 19.

Reed, A. E., Chan, L., & Mikels, J. A. (2014). Meta-analysis of the age-related positivity

    effect: Age differences in preferences for positive over negative information. *Psychology*

    and Aging, 29(1), 1-15.

Roid, G. H., & Koch, C. (2017). Leiter-3: Nonverbal cognitive and neuropsychological

    assessment. In *Handbook of nonverbal assessment* (pp. 127-150). Cham: Springer.

van der Schaaf, M. E., Warmerdam, E., Crone, E. A., & Cools, R. (2011). Distinct linear and

    nonlinear trajectories of reward and punishment reversal learning during development:

    Relevance for dopamine's role in adolescent decision making. *Developmental Cognitive*

    Neuroscience, 1(4), 578-590.

Schmittmann, V. D., van der Maas, Han LJ, & Raijmakers, M. E. (2012). Distinct

    discrimination learning strategies and their relation with spatial memory and attentional

    control in 4-to 14-year-olds. *Journal of Experimental Child Psychology, 111*(4), 644-

    662.

Schwartz, D. L., & Arena, D. (2013). *Measuring what matters most: Choice-based*

    assessments for the digital age. MIT Press.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction,*

    16(4), 475-522.

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-

    constructivist assessments. In T. M. Duffy, & S. Tobias (Eds.), *Constructivist*

    instruction: Success or failure? (1st ed., pp. 34-61). New York: Routledge/Taylor &

    Francis Group.

Shchebetenko, S., Kalugin, A. Y., Mishkevich, A. M., Soto, C. J., & John, O. P. (2020). Measurement invariance and sex and age differences of the Big Five Inventory–2: Evidence from the Russian version. *Assessment, 27*(3), 472-486.

Spielberger, C. D. (1988). *Manual for the state trait anger expression inventory (STAXI)*. Odessa, FL: Psychological Assessment Resources.

Steenhaut, P., Demeyer, I., De Raedt, R., & Rossi, G. (2018). The role of personality in the assessment of subjective and physiological emotional reactivity: A comparison between younger and older adults. *Assessment, 25*(3), 285–301.

Swann, W. B., Jr., & Read, S. J. (1981). Acquiring self-knowledge: The search for feedback that fits. *Journal of Personality and Social Psychology, 41*(6), 1119-1128.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., . . . Wallner-Allen, K. (2013). Cognition assessment using the NIH toolbox. *Neurology, 80*(11 Supplement 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Winstone, N. & Carless, D. (2019). *Designing effective feedback processes in higher education: A learning-focused approach*. London: Routledge.

Wirth, M., & Kunzmann, U. (2018). Age differences in regulating negative emotions via attentional deployment. *Psychology and Aging, 33*(3), 384-398. http://doi.org/10.1037/pag0000254

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson psycho-educational battery*. Chicago: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson psycho-educational battery - Revised*. Chicago: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education, 45*(2), 224–238. https://doi.org/10.1080/02602938.2019.1629390.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH toolbox cognition battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16-33.

Zhuang, Y., Feng, W., & Liao, Y. (2017). Want more? Learn less: Motivation affects adolescents learning from negative feedback. *Frontiers in Psychology*, 8, 76, 1-10. DOI=10.3389/fpsyg.2017.00076.

Zimprich, D., & Mascherek, A. (2012). Measurement invariance and age-related differences of trait anger across the adult lifespan. *Personality and Individual Differences, 52*(3), 334-339.

Zuckerman, M., Brown, R. H., Fischler, G. L., Fox, G. A., Lathin, D. R., & Minasian, A. J. (1979). Determinants of information-seeking behavior. *Journal of Research in Personality, 13*(2), 161-174. https://doi.org/10.1016/0092-6566(79)90028-X